

基于深度强化学习的共享单车调度优化方法

李晓丹^{1,2} 李冉¹ 马迎辉¹

LI Xiaodan LI Ran MA Yinghui

摘要

共享单车因其流动特性,常呈现车辆流动的马太效应,部分区域单车堆积,其他区域却单车匮乏,严重影响车辆周转率,致使单车分布失衡。这种不均衡分布让传统调度方法难以依据实时交通流量、用户需求分布和单车分布等复杂因素动态地选择最优站点,造成调度成本高、效率低。鉴于此,设计一种基于深度强化学习的共享单车调度优化方法。综合步行距离、单车电量和支付费用构建用于评估共享单车调度效果的效用函数。借助构建的效用函数,运用深度强化学习模型在多个调度站点中选择最优站点,并将其作为优化模型的输入。设定用户满意度最大化和调度成本最小化的目标函数,同时考虑供需平衡和调度能力等约束条件来建立调度优化模型。通过迭代算法求解此模型得到 Pareto 最优解集,再用加权求和法根据实际需求选择 Pareto 最优解集中的最优解,生成最优的调度方案。实验结果表明,共享单车调度成本降低幅度随迭代次数增多而增大,证明该方法在降本增效方面的可行性与有效性。

关键词

深度强化学习;共享单车;调度优化;调度站点;用户需求

doi: 10.3969/j.issn.1672-9528.2025.01.045

0 引言

共享单车运营目前面临着诸多挑战。由于共享单车使用灵活,用户需求在时间和空间上分布不均衡,导致共享单车的分布呈现出较强动态性和不确定性。如工作日早晚高峰时段,地铁站、写字楼周边共享单车需求旺盛,而住宅区此时单车过剩;节假日期间,旅游景点共享单车需求则会急剧攀升。这一供需不平衡的状况及单车自身的流动特性,容易引发马太效应,进而对共享单车的周转率产生负面影响。传统调度方法难以充分考虑实时交通流量、用户需求和单车分布等复杂因素,不能动态选择最优调度站点,调度成本高且效率低^[1]。上述因素不仅影响了共享单车运营企业的经济效益,也降低了用户的使用体验。因此,开展共享单车调度优化研究具有重要的现实意义。通过深入研究共享单车的调度问题,探索更加科学、高效的调度优化方法,有助于提高共享单车的运营效率,降低运营成本,提升用户满意度,进而推动共享单车行业的可持续发展^[2]。

鉴于此,相关研究积极探索智能化共享单车调度优化方法以提高调度效率和服务水平。例如,文献[3]构建双目标

优化模型,目标为最小化总需求不满足度与调度成本。考虑高峰时段调度点需求超调度车辆承载能力的情况,设计允许车辆多时段连续调度且可多次访问同一调度点的策略。用非支配排序技术求解,选取最高层级解集为 Pareto 最优解,但缺乏对实时交通流量变化的考虑,非高峰时段交通流量异常时难以调整,对用户需求分布动态变化适应性不足,致使调度成本高、效率低。文献[4]针对大学城共享单车资源分配不均构建单目标数学规划模型,综合调度成本与用户满意度因素。通过调查问卷收集数据、Python 获取实时数据,引入齐次 Markov 过程预测天气、LSTM 预测单车数量变化趋势,用遗传算法求解最优调度路径和数量以提出调度优化方案。但该方法只能预测单车数量变化趋势,不了解单车在不同交通流量区域的分布和用户需求实时变化,难以精准选择最优站点,增加了调度成本。

针对上述方法存在的弊端,本文设计了一种基于深度强化学习的共享单车调度优化方法。深度强化学习模型能够根据实时的交通流量、用户需求分布以及单车的分布等多种复杂因素,动态地选择最优站点。该模型具备出色的适应性,能够更好地应对复杂多变的城市环境。

1 基于效用函数的共享单车调度站点优选

共享单车运营环境复杂,涉及用户、车辆、运营成本等多方面因素。步行距离影响用户获取单车的便利性,单车电量关系到车辆的可用性和调度的可行性,支付费用直接与用

1. 郑州经贸学院大数据与人工智能学院 河南郑州 450000

2. 河南省多模态感知与智能交互技术工程研究中心
河南郑州 450000

[基金项目] 郑州经贸学院青年科研基金项目(QK2302、QK2216)

户满意度和运营收益相关。通过构建包含这些因素的效用函数,能够适应这种复杂的运营环境,对调度效果进行更准确地评估。在构建效用函数后,根据实际的需求和运营情况,运用深度强化学习模型在多个调度站点中,有针对性地选择最优站点,使调度车辆、单车等资源得到更有效地利用。

针对日常通勤中共享单车用户的需求,用户在选择停放及取车站点时主要关注三大因素:步行距离 s_d 、单车平均可用电量 s_f 和骑行费用 s_e 。本研究采用深度强化学习模型,结合实地调研数据校准模型参数,量化用户对不同站点选择的偏好。用户选择某一站点的效用函数可表示为:

$$U(s) = \chi_1 s_d + \chi_2 s_f + \chi_3 s_e \quad (1)$$

式中: χ_1 、 χ_2 、 χ_3 表示深度强化学习模型参数向量。

运用梯度下降法,对深度强化学习模型进行了参数优化,具体标定结果如表1所示。

表1 模型参数标定结果

序号	参数名称	标定值
1	步行最短距离系数	-0.45
2	平均可用电量系数	0.32
3	支付费用系数	-0.28
4	深度神经网络层数	3
5	深度神经网络节点数	128

从表1可以观察到,用户步行最短距离以及支付费用的系数均为负值,这揭示了随着步行距离和费用的攀升,用户倾向于选择这些站点的意愿会减弱;而单车平均可用电量的系数则呈现正值,意味着随着电量储备的增加,用户选择该站点及其单车的可能性会随之增强。这一发现与实际情况高度吻合,据此推导出了用户选择特定取车站点的效用函数,用于量化用户选择特定共享单车调度站点的效用,从而帮助共享单车运营商优化调度,公式为:

$$V(s) = -0.45s_d + 0.32s_f - 0.28s_e \quad (2)$$

2 共享单车调度优化模型构建

共享单车作为城市短途出行的重要方式之一,其分布格局受用户出行习惯影响^[5]。为了精准描绘单车分布状况并优化调度,本文基于效用函数的共享单车调度站点优选结果,构建了调度优化模型。该模型将需求趋势相近的区域划分为虚拟调度“站点”进行协同管理,旨在平衡用户满意度与调度成本两大要素^[6]。在模型构建过程中,设定了调度站点满意度最大化目标函数和调度总成本最小化目标函数,并设定了一系列约束条件,包括单车供需平衡约束和调度能力约束,以确保模型的实用性和有效性。

共享单车调度优化模型构建流程如图1所示。

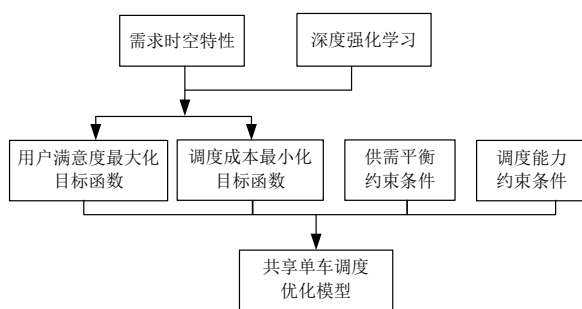


图1 共享单车调度优化模型构建流程

在构建共享单车调度优化模型时,需平衡用户满意度与调度成本两大要素。具体而言,设定了调度站点满意度最大化目标函数,旨在提升用户体验,公式为:

$$\max S = \sum_{j=1}^n s_j \quad (3)$$

式中: S 表示调度站点满意度总和; s_j 表示第 j 个调度站点的满意度; n 表示调度站点的总数。

同时,也制定了调度总成本最小化目标函数,以确保运营效益,公式为:

$$\min C = \sum_{j=1}^n \sum_{k=1}^n c_{jk} p_{jk} \quad (4)$$

式中: C 表示调度总成本; c_{jk} 表示从站点 j 到站点 k 的单位调度成本; p_{jk} 表示从站点 j 到站点 k 的调度量。

在构建共享单车调度优化模型的过程中,必须设定一系列约束条件以确保模型的实用性和有效性。这些约束条件包括,单车供需平衡约束,即每个区域调入的单车数量应等于调出的数量加上该区域新增的需求,公式为:

$$\sum_{j=1}^n p_{jk} - \sum_{q=1}^Q p_{qj} = d_j \quad (5)$$

式中: p_{qj} 表示站点 q 到站点 j 的调度量; d_j 表示站点 j 的单车需求变化量; Q 表示可调出的站点数量。

调度能力约束,即每次调度的单车数量不得超过调度车辆的最大装载量,公式为:

$$0 \leq p_{jk} \leq W \quad \forall_{j,k} \quad (6)$$

式中: W 表示单次调度中所能承载的最大单车数量。此式表示每次调度的单车数量应在0到最大装载量 W 之间。

3 基于迭代算法与加权求和法的模型最优解选取

调度优化问题往往涉及多个目标,而这些目标之间可能存在冲突情况。例如,提高用户满意度的同时可能会导致调度成本的增加。针对此类问题,本文提出采用迭代算法来求解调度优化模型,并获取Pareto最优解集。之后,使用加权求和法,根据实际需求从Pareto最优解集中选择最合适的解,以形成最终的调度方案。这种方法能够综合考虑多个目标,提供多样化的解决方案、适应复杂多变的环境,并解决单一解的局限性、平衡多个冲突目标以及提高决策效率和准确性。

等问题。

首先，初始化共享单车调度优化模型，设置调度站点间的距离矩阵 D_{jk} （其中 D_{jk} 表示点 j 到点 k 的距离），以及深度强化学习的相关参数。在模型中，第一个周期起始点为调度中心，其他周期起始点为上一周期路径最后访问节点。每组调度选择当前点出发最早的调度车，判断其出发时间是否超调度周期时长^[7]。未超时，依约束条件生成候选调度点集，按状态转移概率公式计算各点选择概率：

$$f(D_{jk})=\frac{1}{D_{jk}+\varepsilon}$$

(7)

式中： ε 表示候选调度点数量。

利用深度强化学习确定下一个调度点，依候选点选择概率更新站点和调度车状态，持续至组内无调度点可访问，再进入下一周期。若遍历完所有周期，下一组调度从第一个周期起始点重新开始，直至全部调度完成。判断当前迭代次数是否达最大迭代次数，未达则继续迭代，达则终止程序并输出结果（含 Pareto 最优解集）。由于 Pareto 最优解集并非唯一，其中存在多个解在各自目标上达到最优且互不支配的情况，因此需要进一步处理以找到满足特定需求的最优解。

为选取最优解，给每个目标分配权重，用加权求和法计算解的总体评价 $U(g)$ 。假设有 m 个目标，每个目标的权重为 w_m ，Pareto 最优解集中的某个解为 g ，该解在目标 m 上的值为 $h_m(g)$ ，则总体评价计算公式为：

$$U(g)=\sum_{m=1}w_mh_m(g)$$

(8)

通过计算 Pareto 最优解集中各解的总体评价，找出评价最高的解作为最终最优解，据此生成最优调度方案，用于指导共享单车调度。

4 实验测试分析

4.1 实验区域

以某城市共享单车调度优化实验区域及其毗邻地带为实例，实证所构建的深度强化学习模型与算法的有效性与效率。实验区设 1 个调度中心、8 个调度站点，具体如图 2 所示。调度中心坐标及各站点位置与需求如表 2 所示。

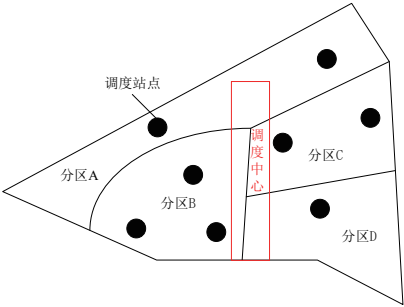


图 2 实验区域分布图

表 2 调度中心及各调度站点位置与需求详情

站点编号	经度 /°E	纬度 /°N	车辆需求 / 辆
调度中心	116.397 128	39.916 527	—
1	116.395 000	39.915 000	50
2	116.396 500	39.917 000	80
3	116.398 000	39.916 000	40
4	116.400 000	39.915 500	60
5	116.401 500	39.917 500	70
6	116.403 000	39.916 500	30
7	116.404 500	39.918 000	90
8	116.406 000	39.917 000	55

在实验区域内，各分区界限分明，调度中心位于各分区几何中心，利于调度指令高效传达。分区 A 为新兴科技园区，骑行需求平稳，调度压力小；分区 B 住宅与商业交织，骑行需求激增，调度调入调出频繁；分区 C 轨道交通网复杂，骑行需求密集，需更多资源应对调度；分区 D 是休闲区域，早晚高峰调度策略需灵活调整以适应出行波动。

4.2 实验准备

实验准备阶段，将实验区域高峰期分为 4 个时段，每时段半小时。早高峰 8:00~10:00，晚高峰 18:00~20:00。调度车辆平均速度依实验城市高峰数据设为 30 km/h。本实验采用深度强化学习模型预测和优化共享单车站点调度需求，其关键参数如表 3 所示。

表 3 深度强化学习模型的关键参数

序号	参数名称	设定值
1	学习率	0.001
2	折扣因子	0.95
3	探索率	0.3
4	神经网络层数	3 层
5	神经元个数	1:128
6	批量大小	64
7	训练轮次	1000 轮
8	优化器	Adam 优化器

将调度计划周期划分为多个等长调度周期，各周期部署一定数量调度车辆。调度初始阶段，车辆空载从调度中心出发，按预设策略到各站点装卸共享单车，站点可被多次访问。调度车辆到站点时，综合考虑下周期单车可用量、预测需求量和自身承载能力确定装卸量。满足调度需求或周期时间耗尽后进入下一周期调度。除首个周期外，各周

期路线起点为上周期最后访问节点，以保证调度连续性与高效性，两辆调度车辆连续两个调度周期的调度示意图如图3所示。

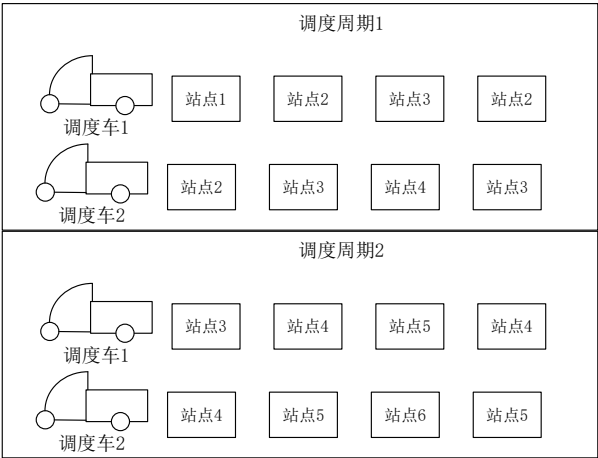


图3 连续两个调度周期内的调度示意图

在调度周期1期间，两辆车分别由调度中心出发，依照各自的最优路径开展调度工作；调度周期2时，这两辆车会依据上一周期的结束位置以及新的调度需求，继续执行新的调度任务。

4.3 实验结果分析

为了验证基于深度强化学习的共享单车调度优化方法的有效性，在保持其他配置不变的情况下，设定不同的迭代次数，并对比调度优化前后的调度成本，结果如表4所示。

表4 调度优化前后的调度成本

单位：元

迭代次数	调度优化前	调度优化后
1	1500	1450
2	1480	1420
3	1460	1390
4	1440	1370
5	1420	1350
6	1400	1330
7	1380	1310
8	1360	1290

从表4中可以看出，调度优化后的成本降低幅度显著大于调度优化前的成本降低幅度，这充分表明基于深度强化学习的共享单车调度优化方法在提高调度效率、降低成本方面具有显著优势。在本文方法中，深度强化学习模型通过不断学习和优化，能够根据实时数据以及构建的效用函数，动态地选择出最优的调度站点。这一操作使得调度方案能够更灵活地适应复杂多变的城市环境，从而显著提高调度效率并降低成本。

5 结语

本文提出的基于深度强化学习的共享单车调度优化方法，在提升共享单车调度效率与服务质量方面成效显著。该方法有效克服传统调度方式在应对单车分布失衡时的不足，通过构建效用函数、设定目标函数并考虑约束条件构建优化模型，再经特定算法求解得到最优调度方案。实验验证其在降低调度成本方面随迭代次数增加效果明显，这表明该方法具有很强的可行性与有效性，为城市交通管理的智能化发展提供了极具价值的范例。展望未来，将进一步深入研究该方法并拓展其应用场景，为城市交通智能化和可持续发展注入更多动力。

参考文献：

[1] 丁云明,陈荔,张昕瑞.基于深度强化学习的柔性作业车间调度问题[J].控制工程,2024,31(7):1185-1194.

[2] 何紫齐,肖磊,李世隆,等.共享单车调度模型和方法综述[J].中国自行车,2022(5):76-79.

[3] 薛晴婉,瞿麦青,彭怀军,等.基于多目标蚁群算法的共享单车调度优化方法[J].交通信息与安全,2024,42(2):124-135.

[4] 孟英豪,王启阳,王柯人,等.基于Markov过程天气预测的共享单车调度优化研究[J].温州大学学报(自然科学版),2024,45(3):30-41.

[5] 赵静.基于深度强化学习的多路径调度模型[J].甘肃科技纵横,2024,53(4):58-65.

[6] 张建同,戴倩楠,丁烨.考虑需求可拆分的共享单车调度优化研究[J].上海管理科学,2023,45(1):119-125.

[7] 陈植元,林泽慧,金嘉栋,等.基于时空聚类预测的共享单车调度优化研究[J].管理工程学报,2022,36(1):146-158.

【作者简介】

李晓丹(1988—)，女，河南洛阳人，硕士，讲师，研究方向：人工智能。

李冉(1990—)，女，河南郑州人，硕士，讲师，研究方向：计算机应用。

马迎辉(1989—)，男，河南郑州人，硕士，讲师，研究方向：深度学习。

(收稿日期：2024-10-15)